

**ANALISIS PENGARUH DATA PENERIMAAN MAHASISWA BARU  
TERHADAP PRESTASI AKADEMIK MAHASISWA  
MENGUNAKAN METODE *NAÏVE BAYES* DAN *DECISION TREE C 4.5*  
(Studi Kasus di Universitas Muhammadiyah Surakarta)**



**SKRIPSI**

Disusun sebagai salah satu syarat menyelesaikan jenjang Strata 1  
pada Program Studi Informatika Fakultas Komunikasi dan Informatika  
Universitas Muhammadiyah Surakarta

**Oleh :**

**TEGUH KARYO UTOMO**

L200110100

**PROGRAM STUDI INFORMATIKA  
FAKULTAS KOMUNIKASI DAN INFORMATIKA  
UNIVERSITAS MUHAMMADYAH SURAKARTA**

**JUNI 2015**

## HALAMAN PERSETUJUAN

Skripsi dengan judul

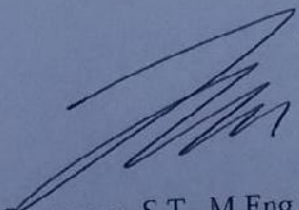
**ANALISIS PENGARUH DATA PENERIMAAN MAHASISWA BARU  
TERHADAP PRESTASI AKADEMIK MAHASISWA  
MENGUNAKAN METODE *NAÏVE BAYES* DAN *DECISION TREE C 4.5*  
(Studi Kasus di Universitas Muhammadiyah Surakarta)**

Ini telah diperiksa, disetujui, dan disahkan pada :

Hari :

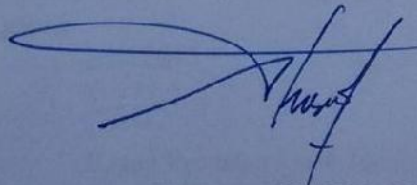
Tanggal:

Pembimbing I



Fajar Suryawan, S.T., M.Eng.Sc. Ph.D  
NIK : 924

Pembimbing II



Yusuf Sulistyo N., S.T., M.Eng.  
NIK : 1197

## HALAMAN PENGESAHAN

# ANALISIS PENGARUH DATA PENERIMAAN MAHASISWA BARU TERHADAP PRESTASI AKADEMIK MAHASISWA MENGUNAKAN METODE *NAÏVE BAYES* DAN *DECISION TREE C 4.5* (Studi Kasus di Universitas Muhammadiyah Surakarta)

Dipersiapkan dan disusun oleh :

**TEGUH KARYO UTOMO**

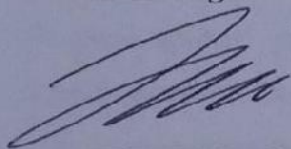
NIM : L200110100

Telah dipertahankan di depan Dewan Penguji

Pada tanggal .....2015

### Susunan Dewan Penguji

Pembimbing I



Fajar Suryawan, S.T., M.Eng.Sc, Ph.D

NIK : 924

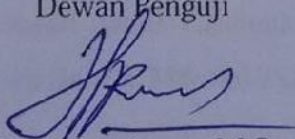
Pembimbing II



Yusuf Sulistyo N., S.T., M.Eng.

NIK : 1197

Dewan Penguji



Dr. Heru Supriyono, M.Sc

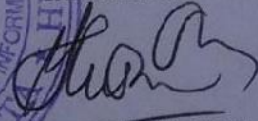
NIK :970

Skripsi ini telah diterima sebagai salah satu persyaratan untuk memperoleh gelar sarjana

Tanggal.....2015



Dekan Fakultas Komunikasi dan Informatika

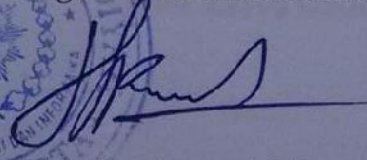


Husni Thamrin, S.T, M.T, Ph.D

NIK : 706



Ketua Program Studi Informatika



Dr. Heru Supriyono, M.Sc

NIK : 970



## DAFTAR KONTRIBUSI


Dengan ini saya menyatakan bahwa skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Berikut saya sampaikan daftar kontribusi dalam penyusunan skripsi :

1. Saya melakukan analisa ini dengan bantuan internet dan buku yang dilampirkan pada daftar pustaka.
2. Program aplikasi yang saya gunakan untuk melakukan penelitian ini adalah Microsoft Excel 2013 dan Rapid Miner 6.
3. Saya menggunakan laptop dengan spesifikasi Core i3-380M, 2.53 GHz, hardik 500 MB dan RAM 2 GB
4. Saya menggunakan data yang telah diberikan oleh Badan Administrasi Akademik melalui bapak Jamil Kasiman dan data PMB dari panitia Penerimaan Mahasiswa Baru UMS melalui mas Fani.

Demikian pernyataan dan daftar kontribusi ini saya buat dengan sejujurnya. Saya bertanggung jawab atas isi dan kebenaran daftar di atas.

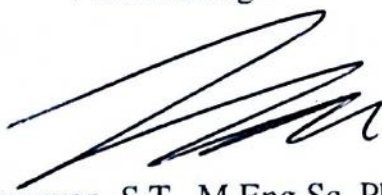
Surakarta, Juni 2015



Teguh Karyo Utomo

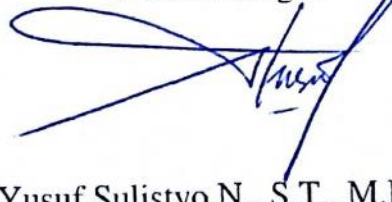
Mengetahui:

Pembimbing I



Fajar Suryawan, S.T., M.Eng.Sc, Ph.D  
NIK : 924

Pembimbing II



Yusuf Sulistyono N., S.T., M.Eng.  
NIK : 1197

## **MOTTO**

“Sesungguhnya sesudah kesulitan ada kemudahan”

( QS. Al-Insyiroh : 5 )

“Dan tolong menolonglah kamu dalam (mengerjakan) kebaikan dan taqwa dan  
jangan tolong menolong dalam mengerjakan dosa dan pelanggaran”

( QS. Al-Maidah: 2 )

“Tiada yang lebih indah selain melakukan hal yang kamu tak senangi namun baik  
bagimu”

( Penulis )

## **PERSEMBAHAN**

1. Kepada Allah SWT yang selalu mendengarkan doa dan memudahkan jalan bagi umatnya.
2. Orang tua tercinta yang selalu memberikan dukungan moral dan materi, kasih sayang, merawat dan mendidik serta senantiasa mendoakan untuk kebahagiaanku dengan penuh keikhlasan.
3. Teman- teman (Kost) Rifki Yakub, Khoerunas, Panggah, Imam dan lain-lain yang selalu membantu menghilangkan rasa stres dan memberikan motivasi.
4. Teman- teman plesiran Labibi, Gunawan, Mardan, Wahyu dan lainnya yang selalu mengajarkan rasa kebersamaan.
5. Teman- teman Informatika angkatan 2011 yang memberikan dukungan dan kebersamaan dalam menyelesaikan tugas akhir ini sehingga selesai dengan baik.

## KATA PENGANTAR

Assalamu'alaikum Wr.Wb

Segala puji syukur penulis panjatkan kehadirat Allah SWT, atas segala rahmat dan hidayahNya yang telah dilimpahkan kepada penulis, sehingga penulis dapat menyelesaikan dan menyusun laporan skripsi ini dengan judul “Analisis Pengaruh Data Penerimaan Mahasiswa Baru Terhadap Prestasi Akademik Mahasiswa Menggunakan Metode *Naïve Bayes* dan *Decission Tree C 4.5* (Studi Kasus di Universitas Muhammadiyah Surakarta Mahasiswa Angkatan 2010)”.

Laporan skripsi ini disusun untuk memenuhi kurikulum pada Program Studi Teknik Informatika Universitas Muhammadiyah Surakarta, sebagai kewajiban mahasiswa dalam rangka menyelesaikan Program Sarjana.

Penulis menyadari bahwa skripsi ini masih jauh dari kesempurnaan, oleh karena itu kritik saran yang membangun dari berbagai pihak sangat penulis harapkan demi perbaikan-perbaikan ke depan.

Terwujutnya skripsi ini tidak lepas dari bantuan berbagai pihak. Oleh karena itu dalam kesempatan ini penulis ingin mengucapkan terima kasih yang sebesar-besarnya kepada :

1. Bapak Husni Thamrin, S.T., M.T., Ph.D. selaku Dekan Fakultas Komunikasi dan Informatika Universitas Muhammadiyah Surakarta.
2. Bapak Dr. Heru Supriyono, S.T., M.Sc. selaku Ketua Jurusan Informatika Universitas Muhammadiyah Surakarta.

3. Bapak Yusuf Sulistyo Nugroho S.T., M.Eng. selaku pembimbing akademik yang telah memberi pengarahan, bimbingan saran serta motivasi selama penulis menimba ilmu di Fakultas Komunikasi dan Informatika Universitas Muhammadiyah Surakarta.
4. Bapak Fajar Suryawan, S.T., M.Eng.Sc, Ph.D dan Yusuf Sulistyo Nugroho S.T., M.Eng. selaku pembimbing yang telah memberikan bimbingan dan pengarahan kepada penulis sehingga dapat menyelesaikan tugas akhir.
5. Segenap dosen dan karyawan prodi Teknik Informatika atas bantuan dan ilmu yang diberikan kepada penulis selama masa perkuliahan hingga dinyatakan mendapat gelar Srata 1.
6. Kepada orang tua yang selalu mendo'akan, memberikan motivasi, semangat dan memberikan yang terbaik kepada penulis.
7. Semua pihak yang tidak dapat penulis sebutkan satu-persatu yang telah membantu hingga terselesaikannya skripsi ini.

Akhirnya penulis berharap semoga skripsi ini berguna bagi semua pihak dan bermanfaat bagi penulis khususnya dan pengguna pada umumnya dalam memudahkan penilaian penentuan peserta sertifikasi guru.

Wassalamu'alaikum Wr.Wb

Surakarta,        Juni 2015

Penulis



## ABSTRAKSI

Universitas Muhammadiyah Surakarta (UMS) merupakan sebuah instansi pendidikan yang besar, berdiri sejak tahun 1958. Pada tahun 2004 UMS melakukan seleksi penerimaan mahasiswa baru (PMB) secara terkomputerisasi dengan sistem *One Day Services* (ODS). Data yang terkumpul setiap tahun cukup banyak sehingga apabila dibiarkan saja maka hanya akan menjadi tumpukan data yang kurang bermanfaat. Oleh karena itu data tersebut akan diolah menggunakan *Data Mining*. *Data Mining* merupakan sebuah metode untuk mencari informasi baru dengan mencari pola pada tumpukan data.

Pada penelitian ini data yang akan diteliti adalah data PMB mahasiswa angkatan 2010 dengan menggunakan algoritma *Naïve Bayes* dan *Decision Tree C 4.5* dan penentuan variabel Y dengan dua cara yaitu sesuai predikat akademik dan kuartalnya. Untuk implementasinya digunakan aplikasi *Rapid Miner 6* dan diharapkan hasilnya bisa digunakan untuk mengetahui sejauh mana pengaruh data PMB dengan prestasi akademik mahasiswa selama kuliah.

Hasil dari penelitian menunjukkan bahwa tingkat akurasi dari kedua metode tersebut berbeda tipis pada variabel Y yang sama yaitu *Naïve Bayes* 66,48% untuk Y berdasarkan predikat dan 33,9%. Sementara untuk *Decision Tree C 4.5* sebesar 66,19% dan 32,4%. Lalu variabel yang paling berpengaruh dalam hasil klasifikasi adalah Jurusan SMA dan Jurusan Kuliah.

**Kata Kunci:** *Data Mining*, *Decision Tree C 4.5*, kuartal, *Naïve Bayes*.

## DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN	ii
HALAMAN PENGESAHAN	iii
DAFTAR KONTRIBUSI	iv
MOTTO	v
PERSEMBAHAN	vi
KATA PENGANTAR	vii
ABSTRAKSI	ix
DAFTAR ISI	x
DAFTAR GAMBAR	xiii
DAFTAR TABEL	xv
<b>BAB I    PENDAHULUAN</b>	<b>1</b>
1.1. Latar Belakang	1
1.2. Rumusan Masalah	3
1.3. Batasan Masalah	3
1.4. Tujuan Penelitian	3
1.5. Manfaat Penelitian	4
1.6. Sistematika Penulisan	5
<b>BAB II    TINJAUAN PUSTAKA</b>	<b>6</b>
2.1. Telaah Penelitian	6
2.2. Landasan Teori	9
2.2.1. <i>Data Mining</i>	9

2.2.2. Tahap-Tahap <i>Data Mining</i>	9`
2.2.3. Metode <i>Data Mining</i>	12
2.2.4. Algoritma C 4.5	13
2.2.5. <i>Naïve Bayes</i>	14
2.2.6. Kuartil	14
2.2.7. <i>Precision</i>	15
2.2.8. <i>Recall</i>	15
2.2.9 <i>Accuracy</i>	15
<b>BAB III METODE PENELITIAN</b>	16
3.1. Metodologi	16
3.1.1 Studi Literatur	16
3.1.2. Pemilihan Objek Penelitian	16
3.1.3. Penentuan Variabel <i>Data Mining</i>	19
3.2. Penentuan Nilai Class	20
3.2.1. Nilai Class Variabel Y	20
3.2.2. Nilai Class Variabel X	21
3.3. Olah Data	22
3.4. Analisis Data	23
3.4.1. Algoritma C 4.5	23
3.4.2. Algoritma <i>Naïve Bayes</i>	24
<b>BAB IV HASIL DAN PEMBAHASAN</b>	26
4.1. Hasil Penelitian	26
4.1.1. Standarisasi Data	26
4.1.2. Pembersihan Data ( <i>Data Cleansing</i> )	27
4.1.3. Integrasi Data	28

4.1.4. Seleksi Data ( <i>Data Selection</i> )	28
4.1.5. Transformasi Data ( <i>Data Transformation</i> )	29
4.1.6. Proses Mining	31
4.1.6.1. Proses Mining dengan Algoritma <i>Decision Tree C 4.5</i> pada <i>Rapid Miner 6</i>	31
4.1.6.2. Proses Mining dengan Algoritma <i>Naïve Bayes</i> pada <i>Rapid Miner 6</i>	42
4.2. Analisa Hasil Penelitian	47
4.2.1. Perbandingan Algoritma <i>Naïve Bayes</i> dan <i>Decision Tree C 4.5</i>	47
4.2.2. Analisa Hasil Klasifikasi pada Data PMB dan Akademik Mahasiswa	49
4.2.2.1. Data dengan Variabel Y Berdasarkan Predikat	49
4.2.2.2. Data dengan Variabel Y Berdasarkan Persebaran kuartal	52
4.2.2.3. Perbandingan Analisa Data dengan Variabel Y berdasarkan Predikat dan Persebaran Kuartal	57
<b>BAB V KESIMPULAN DAN SARAN</b>	59
5.1. Kesimpulan	59
5.2. Saran	60
<b>DAFTAR PUSTAKA</b>	61

## DAFTAR GAMBAR

Gambar 4.1 Rancangan proses untuk <i>Decision Tree C 4.5</i>	32
Gambar 4.2 Hasil pohon keputusan untuk IPK berdasarkan predikat	33
Gambar 4.3 Hasil pohon keputusan untuk IPK berdasarkan kuartal	34
Gambar 4.4 Jurusan SMA sebagai <i>root node</i>	37
Gambar 4.5 Nilai ODS sebagai <i>internal node</i>	40
Gambar 4.6 <i>Leaf node</i>	40
Gambar 4.7 Plot view persebaran nilai IPK berdasar kuartal terhadap Jurusan Kuliah dan Jurusan SMA	41
Gambar 4.8 Rancangan proses untuk algoritma <i>Naïve Bayes</i>	42
Gambar 4.9 Simple Distribution untuk algoritma <i>Naïve Bayes</i>	43
Gambar 4.10 Plot view <i>Naïve Bayes</i> pada variabel Y berdasarkan Predikat	46
Gambar 4.11 Tingkat akurasi, presisi dan recall algoritma <i>Naïve Bayes</i> untuk 4 variabel Y berdasarkan predikat	47
Gambar 4.12 Tingkat akurasi, presisi dan recall algoritma <i>Naïve Bayes</i> untuk variabel Y berdasarkan kuartal	47
Gambar 4.13 Tingkat akurasi, presisi dan recall algoritma <i>Decision Tree C 4.5</i> untuk variabel Y berdasarkan predikat	48
Gambar 4.14 Tingkat akurasi, presisi dan recall algoritma <i>Decision Tree C 4.5</i> untuk variabel Y berdasarkan kuartal	48
Gambar 4.15 Plot view algoritma <i>Naïve Bayes</i> untuk persebaran predikat berdasarkan Jurusan Kuliah dan Jurusan SMA	50
Gambar 4.16 Plot view algoritma <i>Naïve Bayes</i> untuk Jurusan SMA dan	

predikat	51
Gambar 4.17 Plot view algoritma <i>Naïve Bayes</i> pada variabel Y	
berdasarkan persebaran kuartal	53
Gambar 4.18 Plot view persebaran IPK terhadap Jurusan SMA dan	
nilai ODS dengan algoritma <i>Naïve Bayes</i>	54
Gambar 4.19 Plot view dengan aloritma Decisision Tree C 4.5 untuk	
persebaran IPK pada jurusan SMA dan nilai ODS	55
Gambar 4.20 Plot view persebaran IPK terhadap jumlah tes dan	
nilai ODS	56
Gambar 4.21 Plot view dengan algoritma <i>Decision Tree C 4.5</i> untuk	
persebaran IPK terhadap jurusan SMA dan nilai ODS	
berdasarkan predikat	57



## DAFTAR TABEL

Tabel 3.1 Data nilai tes ODS	17
Tabel 3.2 Data riwayat mahasiswa	18
Tabel 3.3 Data prestasi mahasiswa	18
Tabel 3.4 Hasil perhitungan kuartal pada atribut IPK	21
Tabel 3.5 Hasil perhitungan kuartal pada atribut nilai ODS	22
Tabel 4.1 Contoh data hasil normalisasi pada nilai ODS	27
Tabel 4.2 Contoh data hasil kategorisasi jurusan kuliah	27
Tabel 4.3 Contoh data yang telah diseleksi	29
Tabel 4.4 Struktur data yang akan diproses	29
Tabel 4.5 Transformasi data untuk IPK berdasarkan predikat	30
Tabel 4.6 Transformasi data untuk IPK berdasarkan kuartal	30
Tabel 4.7 Jumlah data untuk tiap predikat	35
Tabel 4.8 Jumlah data variabel Jurusan SMA terhadap predikat	35
Tabel 4.9 Nilai information gain untuk tiap variabel X	36
Tabel 4.10 Banyaknya data Jurusan SMA SMA IPS terhadap predikat	37
Tabel 4.11 Banyaknya data Jurusan Kuliah yang berasal dari Jurusan SMA IPS terhadap predikat	38
Tabel 4.12 hasil perhitungan <i>information gain</i> untuk <i>internal node</i>	39
Tabel 4.13 Distribusi probabilitas pada variabel Y berdasarkan predikat	44
Tabel 4.14 Distribusi probabilitas pada variabel Y berdasarkan kuartal	45